ABSTRACT OF THE DISCLOSURE

A method of scheduling in a mixed workload environment. A high priority workload requiring bounded response times is executed on the same system with a low priority workload that is capable of tying up the CPU and multiple volume storage resources of the system by

5   causing multiple concurrent I/O operations, thereby increasing the response times of the high priority workload beyond acceptable bounds. The method of scheduling prevents the response times of the high priority workload from increasing beyond the acceptable bounds by deferring the dispatch of processes servicing the current low priority workload for a time that depends on the priority of the low priority work and by not performing concurrent I/O operations that are

10   requested by the current low priority work, when there is sufficient higher priority activity on the storage resource.